

# Budowa Portfela

## Wprowadzenie

Celem projektu jest zbudowanie w ramach 3-osobowych zespołów modeli pozwalających przewidywać ceny komponentów indeksu WIG30 na najbliższy dzień/dni oraz zbudowanie w oparciu o te przewidywania portfela pozwalającego maksymalizować osiągane zyski. Przewidywany czas trwania projektu to 4 tygodnie. W tym czasie zrealizowane zostać powinny dwa zadania.

1. W ciągu pierwszych dwóch tygodni trwania projektu należy przetestować różne sposoby podejścia do postawionego problemu. Czas ten przeznaczony jest na oczyszczenie danych, próby dopasowania do nich różnych modeli, wyciąganie wniosków z uzyskanych wyników, testowanie jakości uzyskiwanych predykcji i budowanych portfeli. Efektem końcowym powinno być wypracowanie zautomatyzowanej metody doboru instrumentów finansowych do portfela inwestycyjnego.
2. W ciągu kolejnych dwóch tygodni czasu trwania projektu każdy z zespołów przesyła przed rozpoczęciem sesji giełdowej zlecenia kupna i sprzedaży, które w ramach danej sesji chce spróbować zrealizować. Składane zlecenia powinny bazować na sugestiach zaproponowanej uprzednio metody doboru instrumentów finansowych do portfela.

Za rzetelne zrealizowanie projektu uzyskać można łącznie 20 punktów. Przeprowadzenie poprawnych statystycznie i ciekawych analiz w ramach jego pierwszej części daje możliwość zdobycia pierwszych 10 punktów. Kolejne 10 punktów uzyskać można za zwiększenie początkowej wartości portfela o 5% (1 punkt za każde 0.5% wypracowanego zysku).

## Zadanie

### Część 1: Analiza

W trakcie trwania pierwszej części projektu należy skupić się na:

1. oczyszczeniu dostarczonych przez prowadzącego danych wejściowych i przystosowaniu ich do własnych potrzeb,
2. przetestowaniu użyteczności różnych modeli szeregów czasowych do
  - opisu analizowanych danych (spełnienie założeń statystycznych),
  - realizacji zadania (skuteczność w budowaniu zyskownych portfeli inwestycyjnych),
3. zaproponowaniu zautomatyzowanej, finalnej strategii budowania portfela inwestycyjnego.

Poprawna realizacja każdej z wymienionych wyżej czynności będzie miała wpływ na liczbę punktów uzyskanych przez zespół z realizacji tej części projektu.

Budując zautomatyzowane modele statystyczne dla dużych zbiorów danych nie zawsze jesteśmy w stanie zagwarantować spełnienie wszystkich założeń tych modeli dla całego analizowanego zbioru w każdym możliwym przypadku. Z tego powodu nie należy podczas realizacji projektu podchodzić do warsztatu statystycznego nadmiernie restrykcyjnie. Ważna jest jednak świadomość ograniczeń stosowanych modeli. Wiedza o tym, jakie są konsekwencje obchodzenia poszczególnych założeń. Świadomość, których założeń bezwzględnie pomijać nie wolno. Sporadyczne odchodzenie od dobrych praktyk statystycznych motywowane obiecującymi wynikami budowanego algorytmu lub potrzebą uogólnienia metody dla celów automatyzacji będą akceptowane. Dobrą praktyką jest wspomnianie o tego typu problemach w komentarzach do kodu.

Finalny model zaproponowany przez grupę powinien zostać zapisany w postaci pojedynczej funkcji R postaci

```
function(trainData, portfolio, ...) {  
  
}
```

gdzie:

- **trainData** — dane trenujące dostarczone przez prowadzącego (wewnątrz algorytmu mogą być oczyszczane, uzupełniane, modyfikowane, przycinane itp.),
- **portfolio** — obecny stan portfela inwestycyjnego grupy (generowany przez prowadzącego po zakończeniu każdej sesji giełdowej podczas trwania drugiej części projektu),
- ... — wszystkie parametry algorytmu, które mogą podlegać modyfikacji podczas trwania drugiej części projektu.

Funkcja powinna zwracać zestaw transakcji jakie należy przeprowadzić na najbliższej sesji giełdowej w celu maksymalizacji zysku osiągniętego z portfela.

Po zakończeniu pierwszej części projektu grupa powinna oddać do oceny plik R nazwany numerem grupy. Plik ten powinien zawierać wszystkie przeprowadzone przez grupę analizy wraz z komentarzami oraz wspomnianą powyżej funkcję. Plik należy oddać dodając go do odpowiedniego repozytorium GitHub.

## Część 2: Predykcja

W trakcie trwania drugiej części projektu należy skupić się na:

1. terminowym przesyłaniu zleceń składanych przez grupę na daną sesję giełdową,
2. monitorowaniu skuteczności zaproponowanego modelu oraz poprawianiu jego parametrów w razie potrzeby,
3. naprawianiu błędów napotkanych podczas działania algorytmu.

Na tym etapie nie wolno już dokonywać zmian w ogólnej strukturze zaproponowanego algorytmu. Dozwolone jest natomiast dokonywanie jego napraw, jeżeli z jakiegoś powodu nie działa on poprawnie. Można również modyfikować jego parametry, jeżeli wydaje się on słabo współpracować z obecnie zadanymi.

W trakcie trwania tej części projektu:

- po zakończeniu sesji giełdowej, do godziny 20:00, w odpowiednim folderze na Dysku Google umieszczany będzie przez prowadzącego pliki RDS z danymi trenującymi oraz obecnymi stanami portfeli inwestycyjnych wszystkich grup,
- przed rozpoczęciem kolejnej sesji, do godziny 8:00, każda z grup musi w odpowiednim folderze na Dysku Google umieścić pliki RDS ze swoimi zleceniami kupna i sprzedaży na daną sesję.

Każde 0.5% zysku wypracowane przez grupę w trakcie trwania tej części projektu przełoży się na 1 punkt z projektu.

## Role w ramach zespołu

W ramach każdego 3-osobowego zespołu należy wyznaczyć po jednej osobie pełniącej każdą z poniższych ról:

- **analizy** — osoba zajmująca się analizami jest bezpośrednio odpowiedzialna za realizację części pierwszej projektu, ponosi ona wszelkie negatywne konsekwencje nie wywiązania się zespołu z realizacji tej części projektu,

- **raportowanie** — osoba zajmująca się raportowaniem jest bezpośrednio odpowiedzialna za realizację drugiej części projektu, ponosi ona wszelkie negatywne konsekwencje nie wywiązania się zespołu z realizacji tej części projektu,
- **zarządzanie** — osoba zajmująca się zarządzaniem zespołem jest odpowiedzialna za terminowe i rzetelne wykonywanie zadań przez wszystkich członków grupy, odpowiada ona solidarnie z osobą zajmującą się analizami oraz sobą zajmującą się raportowaniem za wszelkie problemy wynikłe w trakcie realizacji zadania.

Główne powody do ukarania jednego z członków grupy odjęciem punktów to:

- nie wywiązywanie się w terminie z powierzonych zadań,
- rażące niedbalstwo w realizacji powierzonej części projektu.

## Terminy

- od 04.01.2016 do 18.01.2016 - realizacja pierwszej części projektu.
- od 20.01.2016 do 02.01.2016 - realizacja drugiej części projektu.

## Dodatki

### Struktura danych trenujących

Dostarczone dane trenujące obejmują okres od 1 stycznia 2012 roku do najbardziej aktualnej, zakończonej sesji giełdowej. Zawierają one ceny zamknięcia wszystkich spółek giełdowych wchodzących w skład indeksu WIG30.

```
readRDS('wig30components.RDS') -> close
head(close[, 1:5], 20)
```

##		acp	alr	att	bhw	bzw
##	2012-01-01	NA	NA	NA	NA	NA
##	2012-01-02	39.321	NA	26.497	53.355	203.99
##	2012-01-03	38.535	NA	26.477	53.787	205.36
##	2012-01-04	38.713	NA	26.497	54.891	205.36
##	2012-01-05	38.470	NA	26.156	54.063	204.45
##	2012-01-06	NA	NA	NA	NA	NA
##	2012-01-07	NA	NA	NA	NA	NA
##	2012-01-08	NA	NA	NA	NA	NA
##	2012-01-09	38.795	NA	25.543	54.497	203.53
##	2012-01-10	38.915	NA	25.611	52.804	205.36
##	2012-01-11	39.143	NA	24.617	53.039	205.36
##	2012-01-12	39.321	NA	25.016	54.733	205.36
##	2012-01-13	39.321	NA	24.938	55.363	203.53
##	2012-01-14	NA	NA	NA	NA	NA
##	2012-01-15	NA	NA	NA	NA	NA
##	2012-01-16	39.321	NA	24.890	57.331	203.53
##	2012-01-17	38.422	NA	25.133	57.962	201.72
##	2012-01-18	37.529	NA	25.104	58.237	203.53
##	2012-01-19	37.108	NA	25.026	56.190	203.63
##	2012-01-20	37.464	NA	24.928	55.521	204.26

Przygotowując dane do analizy należy zwrócić uwagę na to, że:

- w danych uwzględniono dni w których nie odbywają się sesje giełdowe (weekendy, święta),
- niektóre z instrumentów z różnych powodów nie były notowane podczas niektórych sesji giełdowych (np. zawieszenie notowań),
- różne instrumenty finansowe były wprowadzane do obrotu giełdowego w różnym czasie.

## Struktura zlecanych transakcji

Zlecane do wykonania w trakcie sesji transakcje powinny być dostarczane w postaci pliku RDS o poniższej strukturze.

```
data.frame(  
  row.names=c('acp', 'sns', 'pko'),  
  quantity=c(100, -2000, 200),  
  value=c(56.00, 4.00, 27.11)  
)
```

```
##      quantity value  
## acp         100 56.00  
## sns        -2000  4.00  
## pko          200 27.11
```

- Nazwa kupowanego/sprzedawanego instrumentu powinna zostać podana jako nazwa stosownego wiersza ramki danych.
- Liczba kupowanych jednostek powinna zostać podana w kolumnie **quantity** jako wartość dodatnia, natomiast liczba sprzedawanych jednostek jako liczba ujemna.
- Maksymalna cena kupna lub minimalna cena sprzedaży powinny zostać podane w kolumnie **value** w obu przypadkach jako liczby dodatnie.

Planując transakcje kupna należy pamiętać o ograniczeniu jakie stanowi ilość posiadanej gotówki. Planując natomiast transakcje sprzedaży należy pamiętać o tym, aby nie próbować sprzedać większej ilości akcji niż ilość posiadana.

## Struktura portfela inwestycyjnego

Portfel inwestycyjny obrazuje aktualny stan posiadania danej grupy. Zawiera on informacje o ilości posiadanej gotówki (cash) oraz liczbie posiadanych akcji.

```
data.frame(  
  row.names=c(  
    'cash',  
    'acp', 'alr', 'att', 'bhw', 'bzw', 'ccc', 'cdr', 'cps', 'ena', 'eng',  
    'eur', 'gtc', 'ing', 'jsw', 'ker', 'kgh', 'lpp', 'lts', 'lwb', 'mbk',  
    'opl', 'peo', 'pge', 'pgn', 'pkn', 'pko', 'pkp', 'pzu', 'sns', 'tpe'  
  ),  
  quantity=c(1, rep(0, 30)),  
  value=c(  
    50000,  
    56.80, 66.50, 99.50, 71.90, 284.00, 138.55, 22.15, 20.88, 11.30, 12.64,
```

```

48.50, 7.15, 117.15, 10.65, 48.10, 63.49, 5555.05, 27.00, 33.21, 314.00,
6.56, 143.50, 12.79, 5.14, 67.85, 27.33, 68.44, 34.02, 3.81, 2.88
)
)

```

```

##      quantity  value
## cash         1 50000.00
## acp          0   56.80
## alr          0   66.50
## att          0   99.50
## bhw          0   71.90
## bzw          0  284.00
## ccc          0  138.55
## cdr          0   22.15
## cps          0   20.88
## ena          0   11.30
## eng          0   12.64
## eur          0   48.50
## gtc          0    7.15
## ing          0  117.15
## jsw          0   10.65
## ker          0   48.10
## kgh          0   63.49
## lpp          0 5555.05
## lts          0   27.00
## lwb          0   33.21
## mbk          0  314.00
## opl          0    6.56
## peo          0  143.50
## pge          0   12.79
## pgn          0    5.14
## pkn          0   67.85
## pko          0   27.33
## pkp          0   68.44
## pzu          0   34.02
## sns          0    3.81
## tpe          0    2.88

```

## Predykcja w R

Model szeregu czasowego zbudowany na danych od chwili czasu 1 do chwili czasu T może posłużyć do predykcji nieznanej wartości jaką szereg ten przyjmie w chwili  $T + 1$ . W przypadku większości modeli szeregów czasowych w pakiecie R predykcji takiej możemy dokonać wykorzystując funkcję **predict**. Poniżej przedstawiono przykłady zastosowania ten funkcji dla modelu AR(3) oraz GARCH(1,0).

```

na.omit(close$cdr) -> cdr
length(cdr) -> T
log(cdr[-1] / cdr[-T]) -> cdr

```

```

library(TSA)
arima(cdr, c(3, 0, 0)) -> cdrModel
predict(cdrModel, n.ahead=5)

```

```
## $pred
## Time Series:
## Start = 996
## End = 1000
## Frequency = 1
## [1] 0.002993015 0.002554330 0.003458121 0.001286137 0.001310162
##
## $se
## Time Series:
## Start = 996
## End = 1000
## Frequency = 1
## [1] 0.02286358 0.02286438 0.02287402 0.02293444 0.02293449
```

```
library(fGarch)
garchFit(~garch(1,1), cdr, cond.dist='sstd', trace=FALSE) -> cdrGarch
predict(cdrGarch, n.ahead=5)
```

```
##      meanForecast  meanError standardDeviation
## 1  0.001567443 0.02409870      0.02409870
## 2  0.001567443 0.02441650      0.02441650
## 3  0.001567443 0.02463721      0.02463721
## 4  0.001567443 0.02479103      0.02479103
## 5  0.001567443 0.02489847      0.02489847
```